3.0001 DATA DOCUMENTS: A NEW PUBLICATION PLAN FOR SYSTEMATIC ENTOMOLOGY [1]

Ross H. Arnett, Jr. [2, 3]

The availability of the raw data of systematics in the form of journal-published descriptions of new taxa, redescriptions, distribution records, and the records of vouchered specimens, has been delayed in the past by several months to several years from author to user. In the early days of taxonomy, such documents could be distributed only by sailing ships, stage, or on foot, making a few months delay an ordinary occurrence and of little consequence. At that time there were so few taxonomists that the annual research reported even on large groups of organisms could be purchased and stored by all interested persons. A circulation of one or two hundred was the most that could be expected of a journal, and this was enough to pay the printer in those days. A taxonomist of the time found it possible through personal subscription to keep up with the literature in his field, assured that he had missed very little. Moreover, in those unhurried days interests were broad and journal subscribers were interested in and used most of the papers published in each issue. This firmly fixed tradition of publication continues as a procedural ritual without being influenced by the suggestion of changes in the light of modern techniques.

[1] Approved by the Agricultural Experiment Station, Purdue University, Lafayette, Indiana, as Journal Paper no. 3976. Accepted for publication October 20, 1969. Since this was proposed as an Experiment Station Project nearly two years ago, at least four publications have appeared that support this idea: 1) a letter in *Science*, v. 166, no. 3901, pp. 43–44, October 3, 1969, by S. Fred Singer; 2) the SATCOM report published by the National Academy of Science (publication 1717), 1969. This report suggests in recommendation C 12 a publication very similar to the one described here. The generic term for this type of information processing is: Selective Dissemination of Information (SDI). 3) A editorial in *Datamation* (15(12): 183, 1969) shows a history of SDI as early as 1936. 4) Finally, there has appeared as this issue is going to press, an article by E. Yochelson in *Systematic Zoology*, 18: 476–480 which includes a fine discussion of the problems of SDI and the International Code of Zoological Nomenclature, and a proposed solution.

[2] I acknowledge with thanks the helpful suggestions made by Dr. Richard H. Foote, U. S. Department of Agriculture, and the manuscript review committee of the Department of Entomology at Purdue University, including Drs. Ronald L. Giese and Virginia Ferris.

[3] Department of Entomology, Purdue University, Lafayette, Indiana 47907.

The speed up of transportation and communication has been paralleled by an increase in scientific production. The result is that current literature accumulates in nearly disastrous proportions, resulting in an information explosion that alarms every scientist because of the consequential low level of current awareness even among specialists. Even with the speed-up of communication, little has been done to reduce the delay of information transfer from author to user. The publication lag remains the same as it was in sailing ship days. Yet, immediately available are many inexpensive means for the solution of this problem.

In a recent review by Brown *et al.* (1967) a computer-based system is proposed which will enable a subscriber to receive titles, abstracts, and specially selected documents to meet his personal, and perhaps frequently changing needs. The present paper deals with only one aspect of the problem of information communication, the publication of raw taxonomic data. At the same time, it is thought the principles suggested here might apply equally well to other areas of investigation.

In the belief that most taxonomic data reported as the result of original research are of direct interest only to a few specialists, and that these data should be made available almost immediately to this group, the "Data Document" concept is proposed for immediate use in insect taxonomy. By employing the principles already laid down by Brown, and modified slightly to meet the requirements of the International Code of Zoological Nomenclature, I feel certain the system is both functional and logical. Indeed, it is already in operation in some fields of science, even to the extent of utilizing computer storage.

Within a very few years, perhaps even by the end of this decade, information of the type currently found in most scientific journal articles will be stored for instant retrieval in national or international information centers. The Special Committee on Information Storage and Retrieval of the Entomological Society of America is investigating the feasibility of establishing a data center which, if established, could include the storage of information in the manner described herein. The Interuniversity Communication Council (EDUCOM) is deeply involved in the coordination and unification of a project which will result in the eventual change of procedures that will startle and be opposed by the traditional minded taxonomist.

Already I have heard the objection that long papers are needed for promotion. This is, of course, an administrative matter and not a logical argument against the proposed system. But it is of sufficient importance to prevent an easy and rapid change to the system proposed here. It appears that some wholesale revision of administrators' views must be made now and are going to be necessary in the future. The "publish or perish"

mandate is undergoing serious review in many advancing organizations. The acceptance of the *Data Document* concept, or its equivalent by objective administrators, will be an indication of their true concern over the communication problem.

Further comments received lead me to believe that non-taxonomists will be glad to see taxonomic descriptions disappear from the pages of journals, but they do not believe that *their* data should be suppressed! Although this paper concentrates on a system for taxonomy, it is only because I feel competent to suggest a system for this area of study and not others. I feel equally certain that the same system is needed for all areas of entomology. No one can say that one kind of data is of less or greater general interest and usefulness than another kind. In fact, the entire logic behind this proposal is based upon the need to get information to where it is needed when it is needed by the best possible means.

New optional service started

In anticipation of this change, a new publication service is offered called *Data Documents for Systematic Entomology*. This will be available immediately on an optional basis for authors in the two publications, Entomological News and *The Coleopterists' Bulletin*. The necessary procedures for the use of this service are described in this paper. Authors wishing to take advantage of this service may do so simply by indicating this at the time of submitting their papers. The respective editors will then prepare the typescript for *Data Document* processing. They may suggest to others who submit papers to these publications that they use this service, but for the time being, this will be an option selected by the author. Those who do select this service will receive the normal editing services, reviews, and proof of all data to be published in any form.

Scope of publication.—At present, the series of *Data Documents* will be restricted to articles on insect taxonomy, including biological information on insects, or any information treated as a supplement to the taxonomic data. The publication will cover the world fauna and be open to any author, with the provision that the article is acceptable to the editor and reviewers. Every article will be reviewed before acceptance into this system as it would be for traditional publications.

It is not intended that *Data Documents* will replace articles reporting synthesized data, reviews of groups, or biological phenomena information ordinarily published in conventional journal or book form. Works of general use will be published as complete articles. Archieval material should be treated as *Data Documents*. This will include isolated descriptions of

new species not included in a revision and the extensive descriptive and locality data included in revisions.

Restriction of distribution.—It is obvious that the entire system will be destroyed if the distribution of *Data Documents* is not restricted in some manner. There is no way to restrict orders for *Data Documents* except by direct appeal to logic. Copies of these documents should be considered only as something to be used immediately, kept during the tenure of a specific project, and then discarded. Libraries and individuals cannot afford to store them and they should make no attempt to do so. Arrangements will be made for selected depositories as data centers which may take part in the retrieval processing.

Processing of Data Documents.—Authors will need to know format procedures before final typing of articles. They will be required to prepare a list of index terms or descriptors. When a manuscript is received and the author indicates that he wishes it to be treated as a *Data Document*, the editor will check the author's coding of the article. He will then be able to determine the number of copies needed to supply the subscribers to the topics included in the article. Each *Data Document* will be duplicated by a system permitting the production of only the number needed at the time of issue, i.e., required for advance subscription. The production method is chosen according to the number required by these subscribers, and additional copies are produced as needed. Low volume advanced production will be done by xerography. Higher numbers will be produced by other processes. A metal plate will be made and used for illustrations not suited for xerography. No provision will be made for mass distribution of reprints. Authors will be provided with a few copies for records as required by his sponsoring agency or employer.

Available formats.—Each article will be available in three formats. The titles and *Data Document* citation will be published either in the monthly issle of Entomological News or the quarterly issues of *The Coleopterists' Bulletin* as soon as the article is processed. This assures very prompt publication. This citation will include the descriptors and interested persons may place orders according to their selection from these descriptors. In addition, if appropriate, either an informative abstract or an abbreviated article will be issued as soon as possible after the processing. Descriptive abstracts will not be produced. The abbreviated article will contain those portions of the full article deemed immediately useful to a large number of people. For example, keys to genera and species with brief diagnoses and distributional information, might be extracted from an article and published (with the authors permission and galley proof corrections). Thus the greater mass of data will be stored as a *Data Document*.

Price.—Data Documents will be sold at a fixed rate per document, which will include shipping and handling. These may be obtained either directly from the *Center for the Study of Coleoptera* (CSC) at Purdue University for *The Coleopterists' Bulletin* or from the Institute for the Study of Natural Species [4] for articles in ENTOMOLOGICAL NEWS. Advanced subscriptions at somewhat reduced rates will be provided. Subscribers will be charged only for the parts they select at the time of subscribing or according to their change of option which may be submitted at any time (sample subscription forms and details will be supplied upon request). The ordering procedure is indicated on forms provided.

Editing.—Articles will be edited in exactly the same manner as any article submitted to the respective publications. Reviewers will be asked to comment on the article. Changes required by the editors will need to be considered by the authors as with traditional publications methods. After any necessary changes have been made, and the article is accepted, instead of the usual marking for the printer, the editor will prepare a *Data Document* form as a cover. Illustrations will be reduced to the $8\frac{1}{2} \times 11$ format. The title, code words, and any other necessary information, including the document number, will then be prepared for publication in the next issue of the parent publication.

Format.—Articles submitted should conform to the journal format [5] as closely as pratical to avoid any delays in processing. The title should be carefully thought out to indicate an exact description of the contents. As many keys words as possible should be included, and few non-descriptor words. Whenever possible, titles should be limited to 80 characters including spaces so that they may be fully permuted without loss of context by such services as B.A.S.I.C. (Bioscience Information Service, Inc.). An abstract should be prepared containing every descriptor, including all taxa. These abstracts must be limited to 1600 characters, a size selected because of future computer scanning of these abstracts. If all taxa cannot be listed because of these word restrictions, taxa of a higher category should be substituted to indicate the extent of the organisms included in the article.

The title should include the order and family of the insects discussed in the paper. If more than one order and/or several families are included, the title must reflect this by the use of appropriate descriptors. The geographical area covered and the nature of the data presented will serve to restrict the scope of the article.

[4] 550 Elston Road, Lafayette, Indiana 47905.

[5] Articles using this format will appear in the next issue of ENTOMOLOGICAL NEWS.

Articles should be submitted in typewritten form, double spaced, on $8\frac{1}{2} \times 11$ [6] sheets, one side of the page only. In other words, the editorial policy of the respective journals should be followed exactly as before. Authors will be given the same format freedom as previously. However, they are strongly urged to conform with the indentations, centering, and underlining used in the journals because it will not be possible to make these adjustments by the editor's marks. A black, if possible, carbon ribbon on the typewriter is especially desirable. Authors are advised to follow the published format instructions reviewed here before final typing of the article.

The numbering system.—Data Documents of all kinds are numbered to permit instant identification. The system makes no attempt to code according to taxonomic category or taxa because there can be no agreement on this. The system is kept open-ended by the simple method of numbering each article received. These are logged in, and at the time of acceptance, the date recorded on the typescript as is done now. Numbers not appearing in the parent journal refer to papers withdrawn or papers to be published out of sequence. The date of publication is the date of issue of the respective journal, the actual data that the article become available to users. Distinction between the three types of awareness formats is made and is described below.

A. List of *Data Documents.* The complete list of titles and descriptors will be published in the periodical accepting the article. These titles are prefixed by the number 1, followed by a period, and then the document number. The location of deposits of copies will be indicated with the list. From time to time catalogues and indexes may be issued. The 1 will indicate that the title and descriptors are published in the awareness list. For example, this form of indication might appear as follows:

1.0021 Three previously unrecognized New World species of *Oxacis* (Coleoptera: Oedemeridae), by R. H. Arnett, Jr., Department of Entomology, Purdue University, Lafayette, Indiana 47907 (Data Document Center, ISNS). Descriptors [7]: Oxacis; Peru; Trinidad; California; dist.; ills.; keyr.

B. Abstracts. *Data Documents* are available separately as abstracts, either published in the parent journal, if warranted, or placed on file for subscribers or for individual orders. As explained above, they are re-

[6] Some institutions may use a sheet size $8 \times 10\frac{1}{2}$ inches, which is permitted. However, sheets of this size reproduced by xerography will show a black margin, which is somewhat distracting. Sheets larger than $8\frac{1}{2} \times 11$ are not suited to xerography unless a special reducing lens is used.

[7] These abbreviated descriptors are explained below.

stricted to 1,600 characters for ease in storing and for future retrieval by computer scanning. If it is decided that an abbreviated article is to be published by the parent journal, this will be treated separately, as indicated in C below, and will not be considered as an abstract. Abstracts will be given the same document number as the title but they will be prefixed by the number 2. This will inform the user that the abstract, in addition to the title and descriptors, is published in the awareness journal.

C. Complete *Data Document*. Each *Data Document* is given the same number as the title and the abstract, but with the prefix number 3. Even if the entire article is published in the parent journal, this is done. If an abbreviated article is published, the same number will be given to this because exact excerpts will be taken from the complete document. The only change will be to indicate what has been omitted.

It is obviously necessary that each journal have its own series of numbers, so a complete citation must include the name of the journal, volume and page number (for priority purposes), and the name of the document center storing the document. When the system is adopted by other publications, numbers can easily be followed by a periodical number and issue number for short citation.

The Coding system.—The coding of documents is the most difficult procedure for all information storage systems. It must be done carefully, be open-ended, and provide for the matching of search requests both by individuals and by machine. Much might be said about this, but I am discussing this in a separate publication (Arnett, in press, 1970). Users should be warned, however, that no coding principles other than that for zoological nomenclature have been proposed and accepted by a working majority. The system used here may need to be changed at a later date, and time-consuming adjustment made to facilitate retrospective search.

The codes listed below are in addition to geographical locations and the names of taxa included in the title of the article. These code letters are used to describe the contents of an article in the list of documents. Four-letter words are used because of a computerized retrieval program already in operation at Purdue University that involves eight character words, up to ten retrieval words per computer pass. (This can be rewritten to allow for any number of descriptors for retrieval.) By using these four-character words, we are able to combine two concepts as a single request. Details of the system will be described elsewhere.

BIBL—Bibliography of references to taxa.

BIOL—Host information, habitat preferences, and similar observational biological information.

CATA—Catalog of references to taxa.
CSCO—The Center for the Study of Coleoptera, Purdue University.
DESR—Revised descriptions of taxa previously described (subsequent descriptions).
DIST—Distribution of a taxon, including lists of specimens examined.
ILLS—Illustrated.
ISNS—Institute for the Study of Natural Species.
KEYN—New key for identification of taxa.
KEYR—Reference to existing key.
NCOM—New generic assignment of a previously described species.
NGEN—The description of a previously unrecognized genus.
NSPE—The description and validation of a specific name including its generic assignment, designation and deposition of the holotype specimen.
ODRE—Reprint of the original description available for distribution by the retrieval service.
SYNN—New synonymy.
NDSR—New distribution records for the taxon.
TECN—New technique for the treatment of data specimens or observational data described.

Undoubtedly more code terms will be necessary as the system is put into operation and refined.

Method of citation.—Some authors may still feel that a new taxon is not validated unless it appears with a description in the parent journal. Until the matter has full acceptance, an abbreviated description will be published if requested by the author for the purpose of validation. This may take the form of a diagnosis usually found at the beginning of a formal description, similar to the Latin description required by the botanical code. However, once it is determined that validation is made by the stored document alone, the question of method of citation of the species arises. An example of a catalog citation is given here:

Oxacis marianna Arnett, 1970. Data Document 3:0000, Ent. News, 81:00 (p. 0) (ISNS). The number in parentheses after the journal page citation indicates the document page showing where the description started.

Citing these documents in a bibliography also needs explanation. An example of this follows:

Arnett, R. H., Jr., 1970. 1:0000 Three previously unrecognized New World species of Oxacis (Coleoptera: Oedemeridae). Data Document 3.0000, Ent. News, 81:00 (24 pp.) (Data Document Center, ISNS).

In this case, the number of pages after the journal indication shows the total number of document pages on file.

Author's copies.—Reprints, in the journal sense, for the full document are not available. Abbreviated articles may be reprinted as traditional journal articles. Abstracts and complete documents will be supplied in very limited numbers to authors, as explained above, to satisfy the needs of the individual's sponsoring organization. Further copies may be ordered if necessary, but this is discouraged as discussed above.

Personalized subscription service.—The entire concept of *Data Documents* is based upon the limited reproduction of documents and immediate availability of useful material. Although individual sales are possible and provided for, the most efficient method of distribution is through a personalized subscription service, eventually to be computer controlled. Subscribers may select any combination of articles or abstracts according to the code words they select. They are thereby assured of immediate awareness of material needed for their research. To do this, the subscriber must first indicate exactly the taxa and topics of interest to him, and the geographical restrictions, if any, that he wishes. For example, he may wish to subscribe to all articles on North American Coleoptera, and abstracts of all other Coleoptera articles, except for the families Elateridae and Oedemeridae for which he wants all articles. He may desire abstracts for all articles on pollen feeding insects, or some other combination. Each subscriber will have a code number that will indicate his requirements. Articles corresponding to this number will be sent to the subscriber automatically. The cost of the documents supplied will be deducted from his subscription balance. As soon as subscription money is used, a new subscription bill will be issued. In addition, subscribers may purchase coupons to be used for payment for complete documents they may wish in addition to those they automatically receive. Consequently, subscriptions will be based on quantity and not volume or year. Changes in subscription requests may be made at any time without additional charge. However, additional articles ordered, but not previously subscribed to, will be subject to the document fee.

Advantages and disadvantages of Data Documents.—The advantages of the system seem apparent: speed of information dissemination; economy of space required to store entire issues of a publication; economy of production; readily available copies at anytime—never "out-of-print." The system meets the present demands for a solution to the bulging library. The limited but effective circulation also conserves the user's time.

The apparent disadvantages are: high cost of individual copies and resulting lack of private reprint circulation; varying composition, i.e., type-

writer differences, possible format variation and lack of pleasing typographical art; a greater possibility of alteration of master typescript copy so that exact reproduction of each copy is not assured; the elimination of the "browsing" aspect with current journals; the possibility that libraries will demand the full text of each document and thus defeat the space saving feature of the system.

I believe that the apparent disadvantages are greatly outweighed by the advantages. The high cost of individual articles is more than balanced by the reduced expense for journal subscriptions due to the saving of space and reduction of total pages. Even with the disappearance of the publication of raw data, typographical art continues in the synthesis publications which will continue to be widely circulated. The alteration of the master copy can be controlled by the requirement that verified copies be deposited in key information centers, and this is planned. This will assure also that the terms of the International Code of Zoological Nomenclature are met. "Browsing" actually can be enhanced because more time can be spent reading synthesis articles and noting the *Data Documents* in the references cited. Except that the method of presentation of data may be missed, wider coverage of the literature is possible for any individual through the use of the system, and of course, there are other ways to learn how to present data. The matter of libraries subscribing to the entire series of documents may be discouraged by the price of the publications, and by a clear understanding with librarians of the nature of the system.

The future of the system.—The format is designed for easy and eventual automatic data processing. It is not beyond the margins of possibility that all existing systematic entomology data can be gathered, coded, and reprocessed for storage and retrieval by the use of this system. The same principles apply to all other kinds of publications including those with physiological, ecological, and experimental data. Once this is done, there will be no need for the traditional literature search and no need for the complicated rules of nomenclature now so laboriously followed.

Summary

This plan provides an open-ended and flexible system fitted to automatic data processing, awaiting only the increased availability of computer time and a unified processing procedure. The *Data Document* concept is essentially a refinement and wider application of the same system used by Dissertation Abstracts® for theses, except that it eliminates the need for microfilming. I feel that both *Data Documents* and the xerography edition of Dissertation Abstracts meet all of the requirements of publication re-

quired by the Code and that to republish any of this material for the sake of meeting Code requirements is redunant. A thesis, like a *Data Document*, should be prepared for final publication and treated as such when it becomes available in either form. Citations should be made to these documents and priority established on the basis of the date of issue of each.

Data Documents are currently produced by the Center for the Study of Coleoptera (CSC) and the Institute for the Study of Natural Species (ISNS). Other information centers are planning similar publications and services. To be effective, however, all data centers must be linked together (a network) or a chaotic situation will soon result.

LITERATURE CITED

ARNETT, R. H., JR., in press. Entomological Information Retrieval Aids, Institution for the Study of Natural Species.

BROWN, W. S., PIERCE, J. R., and TRAUB, J. F., 1967. The future of scientific journals. Science, 158: 1153–1159.

2.0001 Data Documents, a new publication plan for systematic entomology.

ABSTRACT.—Data Documents is a selective dissimination of information system (SDI) used to control the distribution of information. Documents are stored in a data document center for reproduction and distribution as needed. Awareness of the existence of these documents is published in the parent journal either by title, by title and abstract, or by complete publication. The system is optional, but authors are requested to use the system of storage of data not immediately and generally useful. Descriptors that serve as retrieval codes are provided for each document for automatic data processing. Subscribers may order documents by these descriptors. The system is immediately effective for two journals, ENTOMOLOGICAL NEWS, and the *Coleopterists' Bulletin*.—R. H. ARNETT, JR.

Descriptors: Data Documents; Systematic Entomology; SDI.